

Первичная обработка данных станций космических лучей: алгоритм, вычислительная программа, реализация.

Белов А.В., Блох Я.Л., Клепач Е.Г., Янке В.Г.

*Institute of Terrestrial Magnetism, Ionosphere and Radiowave Propagation (IZMIRAN),
142092, Troitsk, Moscow region, Russia*

Работа посвящена анализу качества работы наземных детекторов космических лучей (нейтронный монитора, мюонный телескоп). Рассмотрены различные типы аппаратурных вариаций, их возможные причины и способы их выделения. Для сравнения однотипных каналов предлагается метод логарифмов отношений, имеющих явные преимущества перед обычным методом отношений. Подробно обсуждается алгоритм первичной обработки данных многоканального детектора, в котором последовательно используется понятие эффективности отдельных каналов и всего детектора.

Приводится ссылка на электронную публикацию программы-редактора, позволяющая определить эффективность каждого канала, выделить в соответствии с заранее выбранным критерием "неисправные" каналы, отредактировать, т.е. исправить данные этих каналов, и кроме того, для получаемых характеристик оценить статистические погрешности как для каждого канала в отдельности, так и для детектора в целом.

Такой контроль и редактирование данных особенно важен при публикации данных детекторов в реальном времени, когда требования к качеству данных максимально высокие.

1. Цели и отличительные особенности первичной обработки данных.

Данные наземных наблюдений космических лучей обрабатываются для того, чтобы, в конечном счете, получить определенные геофизические и астрофизические сведения, прежде всего о функции распределения космических лучей в околоземном космическом пространстве. При этом, как правило, используются данные многих станций и достаточно сложные модели, описывающие первичные космические лучи и их взаимодействие с магнитосферой и атмосферой Земли. Чтобы такая обработка данных (можно называть ее обработкой высокого уровня) была успешной, данные отдельных станций должны быть предварительно проверены и подготовлены, т.е. должны пройти первичную обработку.

Таким образом, главной целью первичной обработки данных космических лучей является подготовка полноценного экспериментального материала для обработки более высокого уровня. Сама по себе первичная обработка не предназначена для извлечения какой-либо физической информации, кроме той, которая имеет самое прямое отношение к работе детектора. Однако именно от качества проведения первичной обработки во многом зависит количество и качество получаемой из данных о космических лучах информации о межпланетном пространстве, магнитосфере и атмосфере Земли.

Первичная обработка данных наземных наблюдений космических лучей обычно проводится на самой станции, непосредственно там, где данные получены. При этом данные каждой станции обрабатываются независимо и используется большой объем вспомогательной информации, необходимость дальнейшего хранения которой после первичной обработки отпадает.

Кроме того, на современном этапе очень важным является проведение первичной обработки и подготовка полноценного экспериментального материала в реальном времени. Это связано с тем, что именно в реальном времени в настоящее время включаюся в ресурсы интернет все большее число экспериментальных данных.

Первичная обработка может в себя включать:

- 1) проверку данных на достоверность, оценку качества работы аппаратуры;
- 2) выявление и исключение аппаратурных вариаций;
- 3) введение поправок на метеорологические эффекты;
- 4) введение поправок на просчеты, случайные совпадения и т.п.
- 5) оценку статистической погрешности данных, прошедших первичную обработку;
- 6) приведение данных к стандартной форме, подготовку их к обмену, публикации, длительному хранению и дальнейшей обработке.

Некоторые из вопросов первичной обработки данных допускают как аппаратурное, так и программное решение. Здесь мы будем обсуждать только программные решения, причем основное внимание будет уделено выявлению и исключению аппаратурных вариаций и оценке экспериментальных данных с помощью контроля эффективности детекторов. Термин "эффективность" детектора (или канала регистрации) интуитивно понятен и часто встречается. Мы также будем пользоваться этим термином, а его определение попробуем дать в одном из следующих параграфов.

2. Развитие методов контроля качества данных в наземных наблюдениях космических лучей.

На первом этапе непрерывных наблюдений основным детектором космического излучения была ионизационная камера [1,2]. Сам принцип действия этого детектора помогал поддерживать его эффективность постоянной. Ток камеры, создаваемый космическим излучением, сравнивался с током, создаваемым внутренним радиоактивным источником. Постоянно измерялась разность этих токов, а для контроля компенсационного тока через равные промежутки времени измерялась сумма токов. Кроме того периодически проводилась калибровка детектора внешним источником ионизирующего излучения.

После появления нейтронных мониторов типа IGY [3], а позднее супермониторов типа NM64 [4], вопрос о контроле их эффективности пытались решать аналогично, периодически проводя калибровку монитора с помощью источника нейтронов. Однако этот метод, в целом, себя не оправдал. Он требовал хранения на станциях радиоактивных источников и дополнительных затрат рабочего времени, нарушал непрерывность данных и не позволял организовать непрерывный контроль эффективности детектора. Состав, энергетический спектр и угловое распределение частиц калибровочного источника совершенно не совпадали с аналогичными характеристиками космического излучения. Эти и некоторые другие недостатки метода калибровки привели к тому, что к настоящему времени от него практически полностью и повсеместно отказались.

Развитие методов контроля качества данных пошло по другому пути. Уже монитор типа IGY делился на две независимые, практически одинаковые секции. Это помогало обеспечить непрерывность данных и позволило контролировать их качество, сравнивая показания секций между собой [5]. Но таким образом можно было только констатировать нормальную работу детектора (если вариации, регистрируемые двумя секциями, совпадали между собой), а отличить неисправную секцию от исправной без привлечения данных других детекторов зачастую было невозможно.

Появление нейтронных мониторов типа NM64, состоящих из трех секций, позволило поднять контроль данных на качественно новый уровень. Здесь сопоставление данных различных однотипных секций уже дает возможность выделить неисправную секцию (но не более одной). Для четырехсекционного детектора не опасен и одновременный выход из строя двух секций.

Заметим, что стандартный мюонный вертикальный телескоп состоит из четырех однотипных секций и для него, с незначительными поправками, подходят те же методы выявления аппаратурных вариаций, что и для нейтронного монитора. Контроль качества данных наклонных мюонных телескопов требует несколько иного подхода.

Таким образом в основе современных методов внутреннего контроля наземных детекторов космических лучей и качества их данных лежит деление детектора на несколько (≥ 3) однотипных, практически одинаковых секций и сравнение данных этих секций между собой. Такой метод (его можно называть секционным) обладает несомненными достоинствами. Он сохраняет

непрерывность данных, дает возможность постоянного контроля и легко поддается автоматизации. Описанные в литературе [6-19] алгоритмы первичной обработки данных станций космических лучей представляют собой различные варианты секционного метода.

Этот метод увеличивает свою эффективность при увеличении числа секций. Самым правильным, по-видимому, является выделение данных каждого нейтронного счетчика в независимый информационный канал. При этом количество подлежащей отбраковке информации уменьшается в несколько раз и соответственно уменьшается статистическая погрешность данных детектора. Не менее важным представляется и следующее обстоятельство. В трех секционном детекторе вероятность того, что аппаратурные вариации, одного знака и близкие по величине, произойдут одновременно в двух секциях достаточно велика. В этом случае исправная секция будет ошибочно признана неисправной и эффективность детектора будет вычислена неверно. Понятно, что вероятностью, связанного с аппаратурными вариациями, статистически значимого одновременного "ухода" в одну сторону, скажем, 10 каналов из 18 можно практически пренебречь.

Говоря про многосекционный принцип регистрации нейтронной компоненты, нужно заметить, что счетчики в нейтронном мониторе не являются одинаковыми, а их показания нельзя считать строго независимыми. Но эти различия и зависимости можно выявить и учесть в процессе контроля работы многоканального детектора.

С увеличением числа информационных каналов существенно увеличивается и количество вспомогательной информации, подлежащей первичной обработке. Но это сейчас уже не может рассматриваться как существенный недостаток. Использование вычислительной техники открывает новые возможности в аппаратурном контроле и заставляет по-новому походить к задачам и методам обработки данных. Нам представляется, что в будущем объем вспомогательной первичной информации будет все более возрастать. Не исключено, что такой информацией станут данные о месте и времени регистрации в детекторе каждой отдельной частицы вместе с амплитудой или даже формой соответствующего импульса.

Пожалуй, единственным существенным недостатком многосекционного метода является то, что он не может выявить изменения эффективности детектора, как целого, т.е. те изменения, которые синхронно происходят во всех каналах. Такого рода изменения не обязательно имеют аппаратурное происхождение. Они могут быть связаны с какими-то переменами, произошедшими поблизости от детектора (например, изменение толщины снежного покрова, перестройка здания и т.п.). Но качество данных детектора они ухудшают точно так же, как и аппаратурные вариации. Поэтому наряду с методами внутреннего контроля, следует развивать также методы, основанные на сопоставлении между собой данных различных детекторов.

3. Типы аппаратурных вариаций.

Все вариации аппаратурного происхождения можно, в первом приближении, разделить на три основных типа: выбросы, скачки и дрейфы. Эти типы вариаций наглядно представлены на рис. 1.

Причиной выброса может быть неисправность источника высокого напряжения, кратковременное отключение питания, наводка, сбой регистратора, сбой при хранении информации на промежуточном носителе и т.п. Выбросы могут быть одиночными или следовать друг за другом в течение короткого времени. Выброс - это практически полностью испорченная информация, подлежащая отбраковке.

Скачок может наблюдаться при замене одного из элементов информационного канала или при резком изменении режима его работы. Как правило, редактирование данных позволяет избавиться от скачка и сохранить полноценную информацию.

Происхождение дрейфа обычно нелегко установить. Возможные его причины: приработка, старение или недостаточная температурная стабильность какого-либо элемента информационного тракта и т.п. Иногда причин дрейфа несколько и их трудно разделить. В качестве особой разновидности дрейфа можно рассматривать периодические и квазипериодические аппаратурные вариации, например, суточные или сезонные. Данные, в которых присутствуют аппаратурные дрейфы можно, в принципе, отредактировать так, чтобы сохранить их полноценность, но это значительно труднее, чем в случае скачка.

Все типы аппаратурных вариаций можно трактовать как изменение эффективности отдельных каналов и всего детектора. Можно дать следующие краткие определения. Выброс - кратковременное значительное изменение эффективности с быстрым возвратом к прежнему значению. Скачок - резкое изменение эффективности на достаточно продолжительное время. Дрейф - постепенное, сравнительно медленное изменение эффективности.

Предлагаемые определения, так же как и само деление аппаратурных вариаций на три типа, в определенной мере условны, но достаточно удобны для практики.

При обсуждении аппаратурных вариаций везде выше речь шла об изменениях скоростей счета, усредненных (или накопленных) за одинаковые промежутки времени. Чаще всего первичной обработке подвергаются среднечасовые данные. Измеряя средние скорости счета, мы имеем дело с реализациями определенного статистического распределения (в простейшем случае пуассоновского). Пока мы говорили о влиянии аппаратурных вариаций на математическое ожидание данного распределения. Однако аппаратурные факторы могут воздействовать и на распределение в целом. Они могут увеличить или уменьшить его дисперсию или каким-либо другим образом существенно повлиять на форму распределения. В определенном смысле можно говорить об аппаратурных вариациях систематического и случайного характера и об аппаратурных вариациях флуктуаций скоростей счета. Особенно важны аппаратурные вариации дисперсии σ_{device} . Их нужно знать для верной оценки статистической погрешности данных детектора. Кроме того вряд ли можно вполне точно определить аппаратурные вариации и корректно отредактировать данные без одновременного определения их дисперсии.

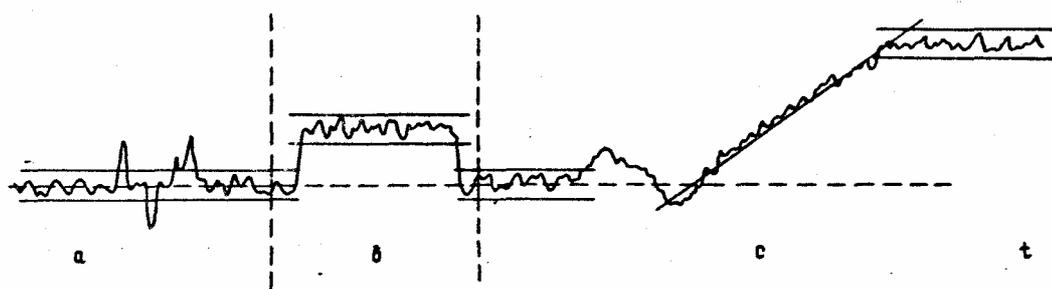


Рис. 1. Классификация аппаратурных вариаций. По оси ординат отложена эффективность канала, а - выбросы, б - скачки, в - дрейф.

Так же как в случае средних скоростей счета можно говорить об аппаратурных выбросах, скачках и дрейфах дисперсии и рассматривать аппаратурные флуктуации как флуктуации эффективности.

Попробуем дать теперь определение самой эффективности.

4. Эффективность детектора.

При определении эффективности следует учесть, что скорость счета детектора зависит не только от количества падающих на детектор частиц, но и от их состава и распределения по энергиям, а также то, что регистрация частиц в детекторе - случайный процесс и эффективность будет, таким образом, случайной функцией многих переменных.

Пусть $S(t) = \sum p_i(t)$ - это сумма вероятностей регистрации, падающих в данный интервал времени на детектор частиц, а $S(t_0)$ - это сумма существовавших в определенный заранее выбранный интервал времени вероятностей регистрации точно таких же частиц. Тогда математическое ожидание отношения

$$\varepsilon(t) = \frac{S(t)}{S(t_0)} = \frac{\sum p_i(t)}{\sum p_i(t_0)} \quad (4.1)$$

можно назвать эффективность детектора в данное время. Эффективность отдельного канала можно определить аналогично.

Возможно и другое определение эффективности, непосредственно говорящее о ее использовании.

Эффективность детектора $\varepsilon(t)$ - это число, на которое нужно разделить наблюдаемую скорость счета $N(t)$, чтобы избавиться от вариаций, связанных с изменениями самого детектора, т.е.

$$\varepsilon(t) = 1 + \delta^{device}(t) \quad (4.2)$$

где $\delta^{device}(t)$ - аппаратные вариации детектора. Таким образом, чтобы избавиться от аппаратных вариаций (отредактировать данные) достаточно знать вариации эффективности детектора:

$$N'(t) = N(t) / \varepsilon(t) \quad (4.3)$$

Эффективность - величина относительная. Поэтому, кроме ее значения в данный момент времени, нужно знать также для какого более раннего момента было справедливо $\varepsilon(t_0) = 1$.

5. Способы сравнения однотипных каналов.

Задачей первичной обработки данных космических лучей является выделение истинных вариаций (т. е. собственно вариаций космических лучей) и освобождение данных от ложных (аппаратурных) вариаций. Однако на первом этапе первичной обработки приходится искать способы добиться обратного: выделить в чистом виде аппаратные вариации, избавившись от вариаций космических лучей. С этой целью при сравнении однотипных каналов чаще всего используются или отношения их скоростей счета

$$r_{ij} = N_i / N_j \quad (5.1)$$

или разности

$$d_{ij} = N_i - N_j \quad (5.2)$$

Пусть вариации отсчитываются от момента времени t_0 в который известны скорости счета N_i^0 и N_j^0 . Тогда для текущего момента времени

$$N_i = N_i^0 (1 + \delta) (1 + \delta^{device}) \quad (5.3)$$

где $\delta = \delta(t)$ - вариация космических лучей, которая предполагается общей для всех каналов, а $\delta^{device} = \delta^{device}(t)$ - аппаратная вариация, различная для разных каналов.

В методе отношений удобнее пользоваться не самими отношениями а следующими величинами (отношениями отношений):

$$r_{ij} = \rho_{ij} / \rho_{ij}^0 = \frac{N_i}{N_j} / \frac{N_i^0}{N_j^0} \quad (5.4)$$

Подставляя в (5.4) выражение (5.3) получим

$$r_{ij} = \frac{1 + \delta_i^{device}}{1 + \delta_j^{device}} = \frac{\varepsilon_i}{\varepsilon_j} \quad (5.5)$$

Таким образом, метод отношений позволяет освободиться от вариаций космических лучей, а отношения отношений простым образом связаны с аппаратурными вариациями и эффективностью каналов. Используя величины r_{ij} и связь $\varepsilon_i = r_{ij} \varepsilon_j$ легко вычислять эффективности. Недостатки метода отношений связаны с нелинейным преобразованием, которому подвергаются скорости счета при вычислении отношений. Вследствие этого изменения эффективностей в каналах i и j сказываются на величинах ρ_{ij} и r_{ij} далеко не одинаковым образом. При 100%-ном уменьшении эффективности (т.е. отключении) канала величина r_{ij} уменьшится на единицу, тогда как отключение канала j приведет к возрастанию r_{ij} до ∞ . Еще важнее то обстоятельство, что распределение флуктуаций отношений ρ_{ij} и r_{ij} сильно асимметрично и далеко от нормального закона. Это не так уж существенно, пока аппаратурные вариации и случайные статистические флуктуации малы. Однако при увеличении аппаратурных вариаций или статистических флуктуаций (например, при переходе к усреднению за малые промежутки времени) такие свойства распределения отношений создают серьезные помехи при статистической обработке данных и ухудшают точность определения аппаратурных вариаций и эффективности.

Использование вместо отношений их вариаций, т.е. величин $1 - r_{ij}$ существенных отличий от рассмотренного случая не имеет и отдельного обсуждения не требует.

Использование разностей в виде (5.2) возможно только для одинаковых каналов, т.е. если $N_i^0 = N_j^0$. На деле приходится использовать приведенные разности

$$d_{ij} = c_i N_i - c_j N_j \quad (5.6)$$

где c_i и c_j - нормировочные коэффициенты, отражающие априорно известные различия каналов в отвечающие условию $c_i N_i^0 - c_j N_j^0 = 0$. Если справедливо (5.3), то с учетом последнего условия

$$d_{ij} = (c_i N_i^0 \delta_i^{device} - c_j N_j^0 \delta_j^{device})(1 + \delta) = 0 \quad (5.7)$$

Мы видим, что выражение (5.7) предельно упростится, если нормировочные коэффициенты выбрать из условий

$$c_i N_i^0 - c_j N_j^0 = 0 \quad (1 \leq i \leq k, 1 \leq j \leq k) \quad (5.8)$$

$$\sum_{i=1}^k c_i N_i^0 = k \quad (5.9)$$

Тогда

$$c_i = \frac{1}{N_i^0} \quad \text{и} \quad d_{ij} = \frac{N_i}{N_i^0} - \frac{N_j}{N_j^0}$$

или

$$d_{ij} = (\delta_i^{device} - \delta_j^{device})(1 + \delta) = (\varepsilon_i - \varepsilon_j)(1 + \delta) \quad (5.10)$$

Однако при любом выборе нормировочных коэффициентов разности d_{ij} (и определяемая с их

помощью эффективности) будут зависеть от величины вариации космических лучей и от наблюдаемой скорости счета детектора. По этой причине использование метода разностей можно рекомендовать только если аппаратурные вариации и статистические флуктуации много больше вариаций космических лучей, или же если от вариаций космических лучей можно заранее достаточно надежно избавиться другим образом. Тогда его преимущества (эквивалентность каналов, симметрия распределения) перед методом отношений весомее, чем его недостатки. Вести контроль эффективности отдельного канала можно также, рассчитывая отношение скорости счета данного канала к сумме исправных из оставшихся каналов. Тогда эффективность канала:

$$\varepsilon_i = \frac{N_i \sum N_i^0}{N_i^0 \sum N_i} \quad (j \neq i, j \neq j_m) \quad (5.11)$$

где j_m - номера неисправных каналов. Это позволяет избавиться от истинных вариаций, но такой способ вряд ли удобен. Может быть важнее всего то, что он позволяет полностью использовать принцип многоканальности. Неясно, например, как можно оценить статистическую точность эффективностей каналов, найденных таким способом. От вариаций космических лучей можно избавиться переходя к отношениям разностей скоростей счета однотипных каналов к их суммам, т.е. величинам:

$$\eta_{ij} = \frac{N_i / N_i^0 - N_j / N_j^0}{N_i / N_i^0 + N_j / N_j^0}$$

При этом

$$\eta_{ij} = \frac{\delta_i^{device} - \delta_j^{device}}{2 + \delta_i^{device} + \delta_j^{device}} = \frac{\varepsilon_i - \varepsilon_j}{\varepsilon_i + \varepsilon_j} \quad (5.12)$$

Несмотря на то, что этот способ довольно громоздкий и имеет определенные неудобства, возможно именно его следовало бы рекомендовать, если бы не нашлось способа, к описанию которого мы сейчас перейдем.

Нам представляется, что самым правильным при сопоставлении однотипных каналов детектора для выявления аппаратурных вариаций будет использование вместо отношений логарифмов отношений, т.е. величин

$$l_{ij} = \ln r_{ij} = \ln N_i - \ln N_j + \ln N_j^0 - \ln N_i^0 \quad (5.13)$$

В этом случае при справедливости (5.3) получим

$$l_{ij} = \ln(1 + \delta_i^{device}) - \ln(1 + \delta_j^{device}) = \ln \varepsilon_i - \ln \varepsilon_{jj}, \quad (5.14)$$

а при малых аппаратурных вариациях ($\delta_i^{device} \ll 1$)

$$l_{ij} = \delta_i^{device} - \delta_j^{device} = \varepsilon_i - \varepsilon_j \quad (5.15)$$

Используя логарифмы отношений мы сохраняем все достоинства метода отношений: избавляемся от вариаций космических лучей, получаем простую связь аппаратурных вариаций с экспериментально измеряемыми величинами, получаем простой способ вычисления эффективностей каналов.

Вместе с тем распределение флуктуаций величин l_{ij} симметрично и существенно ближе к нормальному закону, а скорости счета, аппаратурные вариации и эффективности отдельных каналов входят в выражения для l_{ij} практически одинаковым образом. В связи с этим замена

отношений на их логарифмы должна, как нам кажется, упростить статистический анализ вспомогательной информации, используемой при первичной обработке, и сделать более точным и надежным определение эффективностей и аппаратурных вариаций отдельных каналов и их статистических погрешностей.

Везде выше мы полагали, что аппаратурные помехи приводят к умножению скорости счета на некий фактор (т.е. входят в выражение для скорости счета мультипликативно) и можно пользоваться выражением (5.3). По-видимому, в подавляющем большинстве случаев именно так и будет. Однако можно себе представить аппаратурную помеху, которая входит в выражение для скорости счета аддитивно (шум электронной аппаратуры, радиоактивный фон и т.п.). Тогда вместо (5.3) следует написать

$$N_i = N_i^0(1 + \delta_i) + \delta_i^{device} \quad (5.3a)$$

где δ_i^{device} будет уже аддитивной аппаратурной вариацией. Легко видеть, что в этом случае метод отношений (и логарифмов отношений) не позволит избавиться от истинных вариаций δ . Здесь эффективнее метод разностей. Действительно, если в (5.6) подставить (5.15), получим:

$$d_{ij} = c_j \delta_i^{device} - c_i \delta_j^{device} \quad (5.16)$$

Если пользоваться отношениями скоростей счета, то в присутствии вариаций космических лучей δ , мультипликативных δ_m и аддитивных δ_a аппаратурных вариаций, эффективность можно определить с точностью до фактора $\delta_a \delta / (1 + \delta_m + \delta_a)$.

Подытоживая изложенное, можно сказать следующее. Самый простой и надежный способ выявления аппаратурных вариаций и контроля эффективностей отдельных каналов дает использование логарифмов отношений скоростей счета. Этот способ можно рекомендовать в качестве основного. Для выявления аддитивных аппаратурных вариаций (например, если обнаружится, что поведение эффективностей коррелирует с вариациями космических лучей) следует пользоваться приведенными разностями или отношениями: (разность)/(сумма).

6. Алгоритм контроля многоканального детектора космических лучей.

Предлагаемый здесь алгоритм выявления аппаратурных вариаций и контроля качества данных практически реализован на ст. Киль, Москва, мыс Шмидта. Алгоритм используется для первичной обработки данных 18-ти или 24-ти канального нейтронного монитора и 12-ти канального мюонного телескопа.

Главные отличительные особенности данного алгоритма:

- использование большого числа однотипных (почти одинаковых) каналов;
- непрерывное вычисление эффективностей отдельных каналов и всего детектора;
- 3) использование с этой цепью логарифмов отношений скоростей счета отдельных каналов.

Обрабатываются скорости счета $N_i (i = 1, \dots, k)$, измеряемые в k отдельных каналах за равные промежутки времени. Для конкретности будем говорить о среднечасовых скоростях счета. Обработка каждого часа проводится последовательно и независимо от других. Текущие эффективности ε_i отдельных каналов и эффективность всего детектора вычисляются по отношению к начальным эффективным, существовавшим в определенный более ранний период (начальный период). Начальные эффективности удобнее всего выбирать равными 1. Для начального периода задаются начальные скорости счета N_i^0 , соответствующие начальным эффективным. Кроме начальных эффективностей считаются известными эффективности ε_i^0 , для

периода, непосредственно предшествовавшего обрабатываемому (будем называть их контрольными эффективностями). В качестве оценки логарифма текущей эффективности канала i принимается средняя из оценок, полученных с помощью соотношения (5.14) по исправным из остальных каналов:

$$\ln \varepsilon_i = \sum_j g_{ij} (l_{ij} + \ln \varepsilon_j^0) / \sum g_{ij} \quad (6.1)$$

здесь в далее предполагается суммирование по j при $j \neq i$, $j \neq j_m$, где j_m - номера неисправных каналов, ε_j^0 - контрольная эффективность канала j , а g_{ij} - вес данного члена суммы, зависящий, в первую очередь, от точности определения величины l_{ij} . Большое число каналов позволяет одновременно с аппаратурной вариацией канала оценить и ее статистическую погрешность, которая будет определяться скоростью счета данного канала и качеством работы всех остальных каналов, используемых при расчете данной эффективности, так что

$$S_i^2 = S_{i1}^2 + S_{i2}^2 \quad (6.2)$$

Здесь

$$S_{i1}^2 = \frac{C_k}{C_F N_i^0 (1 + \delta)} \quad (6.3)$$

где в знаменателе стоит скорость счета данного канала в данный момент времени (C_F - коэффициент пересчета), а коэффициент C_k определяет долю генетически связанных импульсов в этой скорости счета и зависит от типа регистрируемой вторичной компоненты космических лучей, от конструкции детектора и от мертвого времени регистрации. Величина S_{i2}^2 определяется из разброса оценок эффективности, полученных по данным различных каналов:

$$S_{i2}^2 = \frac{\sum g_{ij} (l_{ij} + \ln \varepsilon_i^0) - \ln \varepsilon_i \sum g_{ij} (l_{ij} + \ln \varepsilon_j^0)}{(k - m - 1) \sum g_{ij}} \quad (6.4)$$

где m - число неисправных каналов.

В начале обработки считаются исправными все каналы, кроме не работавших в данный час, т.е. тех, для которых $N_j = 0$. Для всех каналов вычисляются оценки логарифма эффективности с помощью (6.1). Далее эти оценки $\ln \varepsilon_i$ сравниваются с контрольными эффективностями и выбирается максимальная по модулю разность и проводится сравнение

$$\max_{i=1, \dots, k} (|\ln \varepsilon_i - \ln \varepsilon_i^0|) > q \sigma_i^0, \quad (6.5)$$

где $q \sigma_i^0$ - величина, зависящая от выбранного критерия достоверности. Если условие (6.5) выполняется, тогда канал j_m , для которого разность логарифмов эффективностей была максимальной, признается неисправным и процедура вычисления эффективностей повторяется с той единственной разницей, что данные канала j_m в сумме выражений (6.1) и (6.2) уже не входят.

Если условие (6.5) не выполняется, т.е. если текущие эффективности в соответствии с

выбранным нами критерием статистически значимо не отличаются от контрольных эффективностей, то вычисленные на данном шаге эффективности ε_i и их погрешности σ_i , считаются окончательными оценками текущих эффективностей и их погрешностей, а все не отбракованные к данному моменту каналы считаются исправными.

Используемая в (6,5) статистическая погрешность σ_i^0 , вообще говоря, различна для различных каналов и меняется со временем вместе со скоростью счета:

$$\sigma_i^0 = \frac{C_k}{C_F(1+\delta)} \left(\frac{1}{N_i^0} + \frac{1}{(k-m)^2} \sum_j \frac{1}{N_j^0} \right) \quad (j \neq i, j = j_m) \quad (6.6)$$

Из-за того, что сомножитель $(1+\delta)$ неизвестен, пока не разделены аппаратные и "истинные" вариации весь анализ работы детектора приходится выполнять дважды каждый час. Вначале в (6.6) полагается $\delta = 0$. После вычисления отредактированной скорости счета всего детектора в (6.6) подставляется реальная величина $(1+\delta)$ и весь анализ проводится уже с уточненными σ_i^0 .

Качество работы алгоритма зависит от выбора критерия достоверности. Если величина q чересчур мала, часто будет отбраковываться полезная информация. Если же, наоборот, величина q завышена, велика вероятность того, что в определении эффективностей наряду с исправными участвуют и неисправные каналы. При этом неточно будут определяться не только эффективности (и аппаратные вариации) каналов, но и скорость счета всего детектора. Нами было принято значение $q=3$.

Весовые множители g_{ij} можно определить, зная дисперсии отношений скоростей счета каналов i и j . Проще всего это сделать, анализируя разброс величин $(\ln N_i - \ln N_j)$ в течение достаточно

длительного времени (предполагается, что в это время контрольные эффективности ε_i^0 и ε_j^0 не менялись). Вводить весовые множители имеет смысл, если дисперсии отношений удалось получить с хорошей точностью и если достоверно известно, что каналы детектора существенно отличаются

друг от друга. Во многих случаях вполне уместно считать, что все $g_{ij} = 1$.

После того как каналы разделены на исправные и неисправные, скорости счета неисправных каналов заменяются на отредактированные скорости счета

$$N_i^* = N_i / \varepsilon_i \quad (6.7)$$

Особый случай возникает, когда канал i совсем не работает и $N_i = 0$. Тогда и вычисленная эффективность данного канала $\varepsilon_i = 0$ и вместо (6.6) для редактирования данных придется использовать выражение

$$N_i^* = N_i^0 \frac{\sum_j g_{ij} \frac{1}{\varepsilon_j^0} \frac{N_j}{N_j^0}}{\sum_j g_{ij}} \quad (j \neq i, j \neq j_m) \quad (6.8)$$

Если $\varepsilon_i = 0$ выражения (6.6 и 6.7) эквивалентны.

Можно рассчитать несколько суммарных скоростей счета, характеризующих работу всего детектора. Если сложить данные всех каналов, ничего в них не меняя, получится фактическая скорость счета:

$$N = \sum_{i=1}^k N_i \quad (6.9)$$

Сумма отредактированных скоростей счета каналов даёт отредактированную (приведенную к начальной эффективности) скорость счета:

$$N^* = \sum_{i=1}^k N_i / \varepsilon_i \quad (6.10)$$

Сравнение фактической и отредактированной скорости счета позволяет легко определить текущую эффективность всего детектора

$$\varepsilon = N/N^* \quad (6.11)$$

Если в данный час один или несколько каналов детектора были признаны неисправными, для определения приведенной скорости счета N^* была использована не вся фактическая скорость счета N , а только ее, так сказать, "полезная" часть

$$\bar{N}^* = \sum_{i=1}^k N_i \quad (i \neq j_{mm}) \quad (6.12)$$

Отношение

$$P = \bar{N}/N \quad (6.13)$$

можно использовать в качестве "коэффициента полезного действия" детектора. При этом, однако, не надо забывать, что коэффициент будет определяться не только качеством работы самой аппаратуры, но и свойствами выданного алгоритма контроля аппаратуры и редактирования данных. Если для одного часа почти все равно, знаем мы число исправных каналов $k-m$ или коэффициент P , то при усреднении данных за продолжительное время (сутки, месяца, год) для оценки работы детектора намного удобнее использовать характеристику типа (6.12).

Еще более важной характеристикой работы детектора является статистическая погрешность ее скорости счета. Ожидаемую статистическую погрешность можно оценить с помощью той же суммы N^*

$$S_1^2 = \frac{C_k^2}{C_F N^*} \quad (6.14)$$

где C_k^2 - определяется долей связанных между собой импульсов в скорости счета детектора. Ту же статистическую погрешность можно оценить, и, используя многоканальность детектора, по разбросу оценок скорости счета детектора, полученных с помощью каждого из исправных каналов

$$N_i^* = \varepsilon_i^0 \frac{N_i}{N_i^0} \sum_{i \neq j_m} \frac{N_i^0}{\varepsilon_i^0} \quad (6.15)$$

Если весовые коэффициенты каналов $g_{ij} = 1$, то

$$S_2^2 = \frac{1}{m(m-1)} \sum_{i \neq j_m} \left(1 - \frac{N_i^0}{\varepsilon_0}\right)^2 = \frac{1}{m-1} \left[\frac{1}{m(N^*)^2} \sum_{i \neq j_m} (N_i^*)^2 - 1 \right] \quad (6.16)$$

Если по случайным причинам, в данный час $S_2 < S_1$ за оценку статистической погрешности следует взять S_1 .

7. Программа - редактор SuperEditor.

Программу "SuperEditor" можно взять на сервере ИЗМИРАН

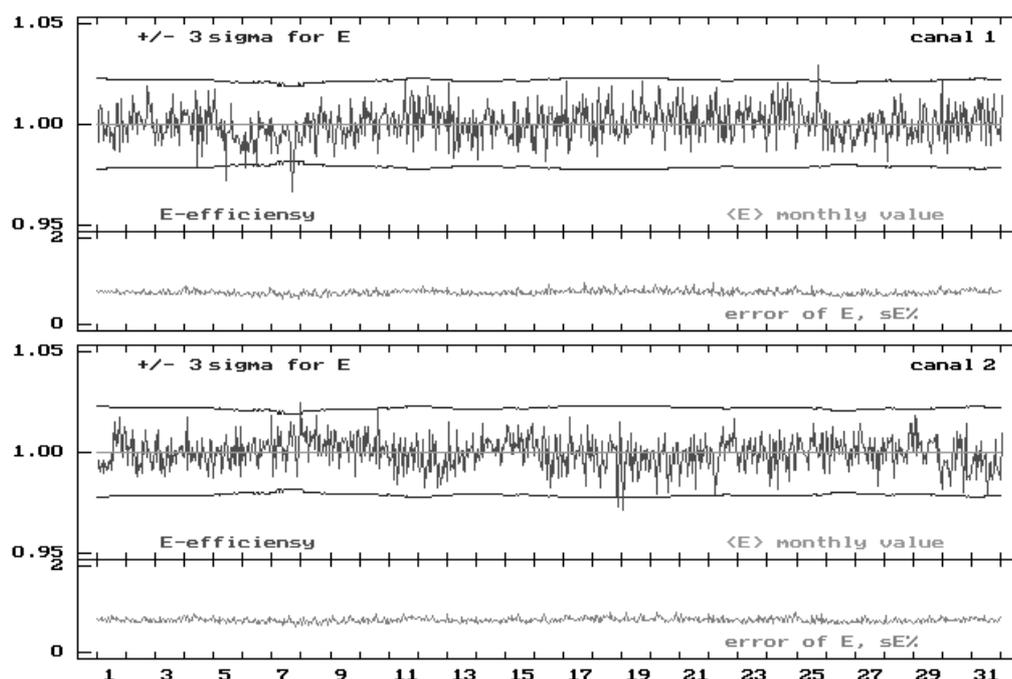
Подпрограмма предназначена для обработки по описанному выше алгоритму данных детектора космических лучей разделенного на однотипные каналы.

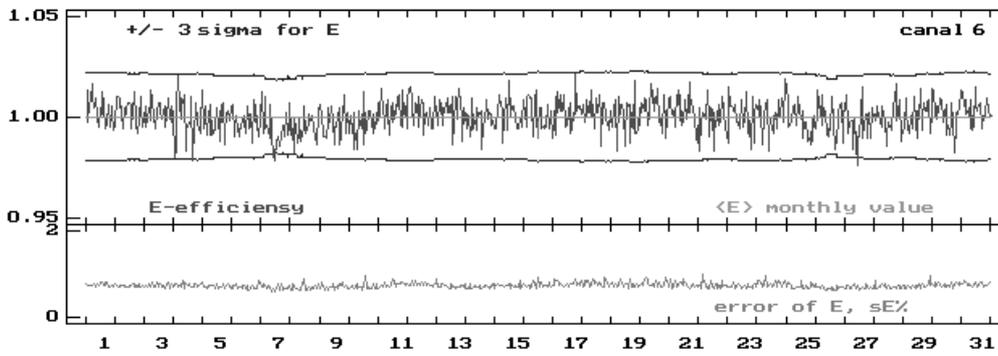
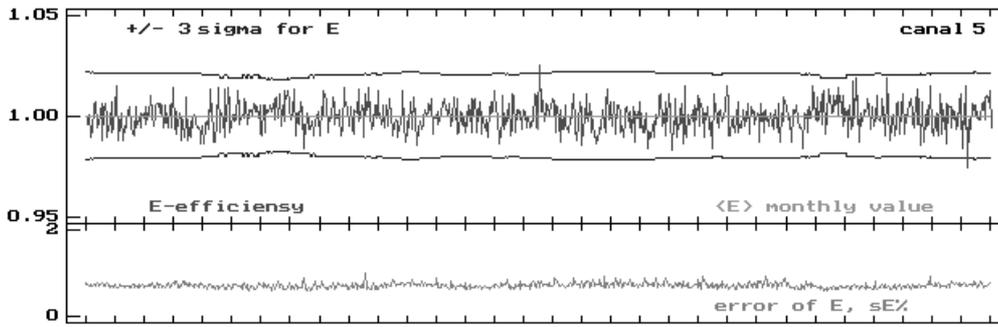
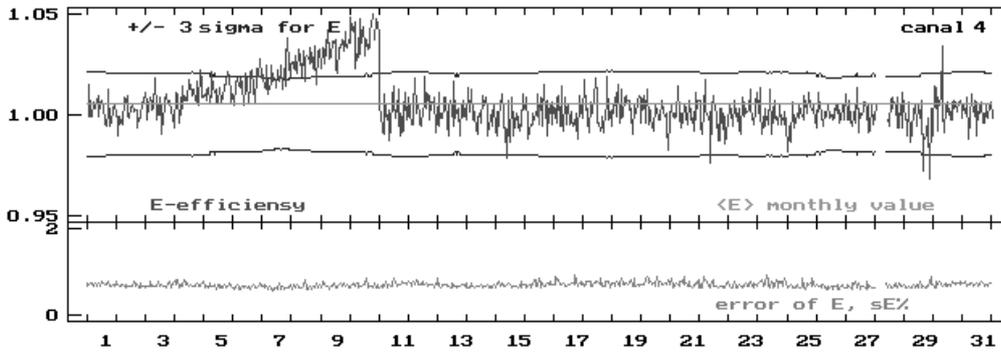
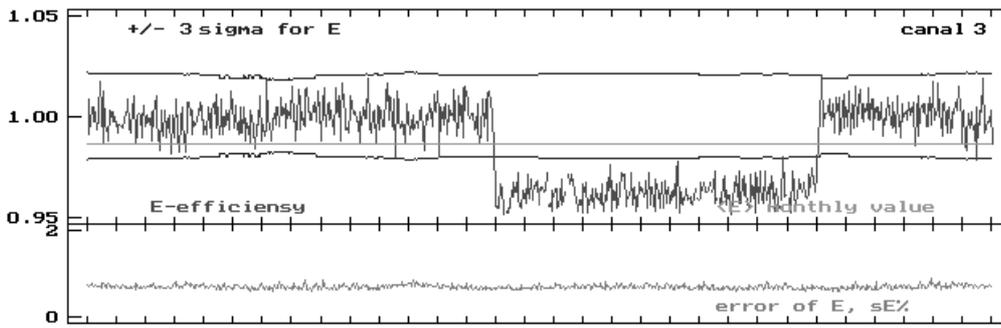
Подпрограмму можно вызвать с помощью оператора:
SuperEditor(K,N0,E0,N,N1,E,SB,SH,S,NS,SN,ST,ES,P,IN,SF,SK,Q)

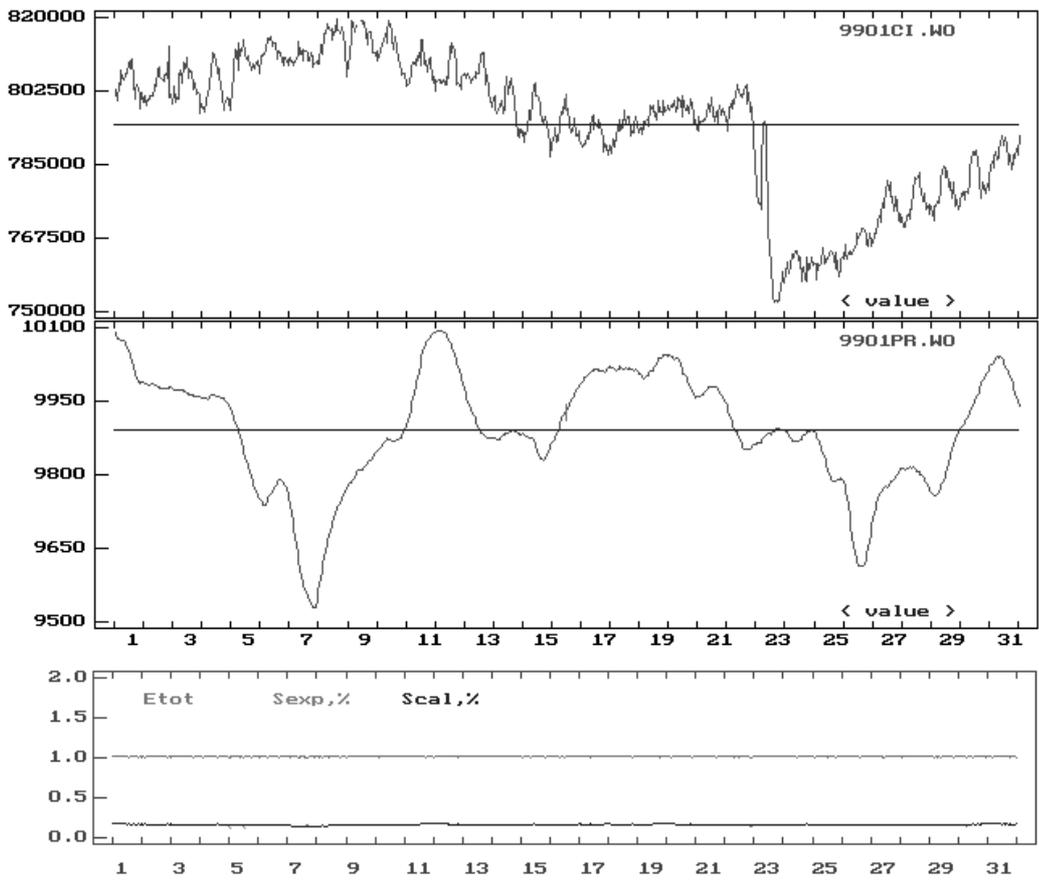
Среди переменных в обращении входными параметрами являются:

к	- число каналов;
N0	- вещественный вектор начальных скоростей счета (этот и все другие входящие в список формальных переменных вектора имеют размерность к);
E0	- вектор контрольных эффективностей каналов;
N	- вещественный вектор обрабатываемых скоростей счета;
SF	- коэффициент пересчета;
CK	- коэффициент, показывающий во сколько раз в отсутствие аппаратурных вариаций дисперсия скоростей счета должна превышать пуассоновскую дисперсию;
Q	- количество ожидаемых стандартных отклонений, при превышении которого канал признается неисправным.
E	- вектор рассчитанных эффективностей каналов (для неисправных каналов эффективность выводится со знаком минус
S	- вектор стандартных среднеквадратичных погрешностей эффективностей каналов;
N1	- вектор отредактированных скоростей счета каналов;
SH,SB	- вектора нижней и верхней границ доверительного интервала, в котором может изменяться эффективность признаваемого исправным каналом;
NS	- отредактированная скорость счета детектора;
SN	стандартная среднеквадратичная относительная погрешность отредактированной скорости счета детектора;
ST	ее теоретическая оценка;
ES	- вычисленная эффективность детектора;
P	- "коэффициент полезного действия" детектора;
IN	- целочисленный вектор-индикатор исправности каналов, составляющие которого после выполнения подпрограммы равны 1- для исправных и 0 - для неисправных каналов.

Если число исправных каналов окажется меньшим $k/2$ или меньшим 2, то эта экспозиция бракуется полностью. Во всех остальных случаях вычисляются следующие выходные параметры:







8. Некоторые результаты.

Программа прошла тестовые испытания, в которых имитировались различные аппаратурные вариации и изменялось число обрабатываемых каналов.

Обсудим применение данной подпрограммы на примере обработки данных нейтронного монитора ст. Москва 24НМ-64; для компактности ограничимся одной секцией, т.е. 6 каналами. Для иллюстрации выберем период с 1 по 31 января 1999; в 3-й и 4-й каналы искусственно введены аппаратурные вариации. На рис. 1 в поведении вычисленных эффективностей каналов можно видеть аппаратурные вариации всех обсуждавшихся выше типов. Пропуски в данных: например, 28 января в 4 канале. Отрицательные выбросы: например, в 4 канале 29 января в 14 и 20 часов. Положительные выбросы: также в 4 канале 30 января. В 3 канале с 15 по 25 января искусственно введен скачек в скорость счета (1.03), что и наблюдается в эффективности. Наиболее ясно выражен аппаратурный дрейф, наблюдаемый в 4 канале начиная с 3 по 10 января (искусственно введен дрейф в скорость счета от 1 до 0.96). На части рассматриваемого периода эффективности 4 канала (6,7 января) существенно отличаются от соответствующих контрольных эффективностей. Однако эти отклонения в основном недостаточно велики, чтобы данные каналы признавались неисправными и большую часть времени отредактированная скорость счета детектора складывается из скоростей счета всех каналов, в том числе и 4 канала. Это приводит к тому, что дисперсия скорости счета детектора, 6 и 7 января, существенно превышает ожидаемую дисперсию. То обстоятельство, что 4 канал (6 и 7 января) в отдельные часы отбраковываются, а в другие часы тех же суток признаются исправными, создает дополнительные флуктуации скорости счета детектора и ее дисперсия может оказаться больше приведенных здесь оценок.

С другой стороны, в каналах 1,2,5,6 кроме пропусков в данных и нескольких кратковременных выбросов, других значительных аппаратурных вариаций не наблюдается. Распределение флуктуаций эффективностей этих каналов близко к нормальному и; согласуется с ожидаемыми значениями среднеквадратичных отклонений (для часового значения эффективности канала $\sim 0,6\%$).

Замена (на основе полученных данных) контрольных эффективностей в 3 канале значительно

уменьшает количество отбракованной информации. При этом существенно уменьшается дисперсия скорости счета детектора, увеличивается коэффициент его использования, уменьшаются флуктуации эффективности и скорости счета детектора.

На рис. 2 исправленная скорость счета монитора и его эффективность, вычисленная по формуле (6.11), а также статистическую погрешность скорости счета - теоретическая оценка (6.14) и экспериментальная (6.16), которые практически совпадают.

9. Возможности совершенствования первичной обработки данных.

Обсудим достоинство и недостатки предложенного здесь алгоритма аппаратного контроля и редакции данных детектора космических лучей.

Главные достоинства данного алгоритма - это, на наш взгляд, его относительная простота, и последовательное использование в нем принципа многоканальности и понятия эффективности. Использование логарифмов отношений скоростей счета каналов делает определение эффективностей и аппаратных вариаций каналов достаточно точным и надежным, даже когда эти аппаратные вариации или статистические флуктуации данных велики. В соответствии с заранее выбранным критерием, алгоритм позволяет определить неисправные в данный момент каналы детектора. Он дает возможность постоянного определения текущих эффективностей каждого канала и всего детектора. Он позволяет получить отредактированную скорость счета детектора, на которую не влияют данные неисправных каналов. Для текущих эффективностей каналов и отредактированной скорости счета детектора алгоритм дает их статистические погрешности, основанные на разбросе данных отдельных исправных каналов. Информация, получаемая с его помощью, позволяет оценивать качество работы детектора и эффективность самого алгоритма.

Все это дает возможность рассматривать данный алгоритм в качестве хорошей основы для первичной обработки данных любого многоканального детектора космических лучей. Вместе с тем достаточно очевидны и недостатки обсуждаемого алгоритма.

Если аппаратная вариация в канале велика, но недостаточно велика, чтобы канал был признан неисправным, он начинает оказывать нежелательное влияние на вычисляемые эффективности других каналов. Одним из возможных способов избавления от этого недостатка может быть переход от средних арифметических по исправным каналам к медленным значениям.

Другие недостатки алгоритма:

- Не может считаться алгоритмом полностью автоматизированной первичной обработки, поскольку время от времени возникает необходимость замены контрольных эффективностей;

- Не вполне приспособлен для выявления аппаратных скачков и дрейфов. Действительно малые скачки и дрейфы данный алгоритм не замечает, а большие, фактически, приравнивает к выбросам. При этом часто отбраковывается полезная информация;

- Не предназначен для выявления аппаратных вариаций дисперсии и более высоких моментов статистического распределения скоростей счета.

Следующим шагом в совершенствовании алгоритма первичной обработки данных должен стать анализ временного хода найденных эффективностей каналов. Это позволит выявить аппаратные скачки и дрейфы в отдельных каналах, своевременно уточнить контрольные эффективности и, в конечном счете, сохранить полезную информацию и увеличить точность данных детектора.

При анализе временного хода эффективностей возможны два принципиально различных подхода: ретроспективный и адаптивный. Первый предполагает, что вначале будут получены оценки эффективностей, затем будет проведена аппроксимация временного хода эффективностей и на ее основе заново отредактированы скорости счета. При этом при редактировании данного часа используется информация, полученная и до и после него. Естественно, такой подход нельзя использовать, если вести первичную обработку в режиме реального времени.

Именно для работы в таком режиме предназначен адаптивный подход. Такой подход, за счет некоторых потерь в точности, позволяет полностью автоматизировать первичную обработку данных и предоставляет самые широкие возможности в оперативном контроле аппаратуры и

быстром извлечении полезной информации.

Другая возможность улучшения методики первичной обработки связана с анализом флуктуаций данных отдельных каналов и с выявлением корреляционных связей между ними. Выше мы предполагали, что данные отдельных каналов независимы, а дисперсия изменяется вместе со скоростью счета во всех каналах одинаковым образом. И то, и другое -очевидные упрощения. Оперативный контроль распределения флуктуаций в каналах и корреляционных связей между каналами позволит выявлять аппаратурные вариации дисперсии, уточнить критерии отбора исправных каналов, точнее- определять аппаратурные вариации, общие для нескольких каналов и т.д.

Одновременно используя два способа сравнения каналов (с помощью логарифмов отношений и приведенных разностей) можно ставить задачу разделения мультипликативных и аддитивных аппаратурных вариаций.

Acknowledgements. This work is supported by Russian Federal Program "Astronomy" and the Russian Foundation for Fundamental Research, grant №. 98-02-17315 and № 99-02-18003.

ЛИТЕРАТУРА

1. Compton A.H., Wollan E.O., Bennet R.D. 'A precision recording cosmic ray meter. - Rev. Scient Instrum., 1934, v. 5, N 12, p. 415-422.
 2. Шафер Ю.Г. Непрерывная регистрация вариаций интенсивности космических лучей ионизационной камерой с автоматическим управлением. - Изд-во АН СССР, Труды ЯФАН, Сер. физ., вып. 2, 1958, с. 7-22.
 3. Simpson I.A. Neutron monitor. - Chicago, 1955, 60 p.
 4. Carmichael H., NM-64 Neutron monitor specification. U.S.A. Deep River Laboratory, 1962, 70 p.
 5. Яноши Л. Теория и практика обработки результатов измерений. - М.: Мир, 1968, 462 с.
 6. Блох Я.Л., Пименов И.А. Стандартный алгоритм для станции космических лучей по проверке постоянства уровня регистрации. - Геомагнетизм и аэрономия, 1967, г. 7, № 5, с. 931-933.
 7. Блох Я.Л., Пименов И.А. Стандартный алгоритм для станции космических лучей. - Геомагнетизм и аэрономия, 1969, т. 9, N 1, с. 53-55.
 8. Блох Я.Л., Дорман Л.И., Пименов И.А. Стандартный алгоритм для станции космических лучей. - В кн.: Космические лучи, М.: Наука, 1969, № 11, с. 149-157.
 9. Дорман Л.И., Пименов И.А., Сацук В. В. Математическое обеспечение исследований геофизических закономерностей на примере космических лучей. - М.: Наука, 1978, с. 152.
 10. Блох Я.Л., Иноземцева О.И., Окулов Ю.И. Алгоритм первичной машинной обработки данных станций космических лучей. - В кн.: Космические лучи, М.: Наука, 1974, № 14, с. 174 - 186.
 11. Вертлиб А. Б., Лузов А. А. К вопросу об оценке достоверности данных по вариациям интенсивности космических лучей. - В кн.: Исследования по геомагнетизму, аэрономии и физике Солнца. - М.: Наука, 1974, вып. 31, с. 161-165.
- Практический метод оценки достоверности данных регистрации космических пучей/Верглиб А.Б., Лузов А.А., Сергеев А. В. и др. - Исследования по геомагнетизму, аэрономии и физике Солнца/. - М.: Наука, 1975, т. 35, с. 161-166.
13. Dyring E., Sporre B. Statical reliability tests of cosmic ray intensity data, Acta Physica Academiæ Scientiarum Hungaricæ, . .Suppl. 2.— 1970, p.. 95-98.
 14. Кудела К. Некоторые критерии качества измерений многоканального нейтронного супермонитора. - Изв. АН СССР, Сер. фаз., 1974, т. 38, № 9, 1982-1985.
 15. Блох Я.Л., Симсарьян Р.А., Старков Ф.А., Типикин А.П. Первичная обработка данных станций космических лучей в реальном масштабе времени на ЭВМ типа "Наири*". - В кн. Космические лучи. - М.: Наука, 1975, Ns 15, с. 203-206.
 16. Блох Я.Л., Позняк А.С., Симсарьян Р.А. Некоторые принципы и алгоритмы построения систем

- первичной обработки данных СКЛ. -В кн.: Космические лучи. - М.: Наука, 1976, N 16, с. 45-50.
- 17.Блох Я.Л., Позняк А.С., Симсарьян Р.А. Адаптивный алгоритм обнаружения аппаратурных сбоев при обработке данных СКЛ. В кн.: Космические лучи, М.: Наука, 1976, № 16, с. 5-1-55.
- Блох Я. Б., Либин И.Я., Симсарьян Р.А., Старков Ф.А. Алгоритм обнаружения аппаратурных сбоев данных супермониторов. В кн.: Космические лучи - М.: Радио и связь, 1983, № 22, с. 56-61.
- Belov A., Blokh Ya., Klepach E., Yanke V.Primary Processing of Cosmic Ray Station Data: Algorithm, Computer Program and Realization. сб. Космические лучи, N25, 113-134, М., Наука, 1988.